

Proteomics: Principles and Techniques - Video course

COURSE OUTLINE

An introduction to proteomics: Basics of protein structure and function, An overview of systems biology, Evolution from protein chemistry to proteomics; Abundance-based proteomics: Sample preparation and prefractionation steps, Gel-based proteomics - two-dimensional gel electrophoresis (2-DE), two-dimensional fluorescence difference in-gel electrophoresis (DIGE), Staining techniques.

Quantitative proteomics - stable isotope labeling by amino acids in cell culture (SILAC), isotope-coded affinity tag (ICAT), isobaric tagging for relative and absolute quantitation (iTRAQ); Central role of mass spectrometry: ionization sources, mass analyzers, different types of mass spectrometers; Functional proteomics: Recombinational cloning, Interactomics - techniques to study protein-protein interactions, yeast two-hybrid, immunoprecipitation, protein microarrays, Nucleic Acid Programmable Protein Array (NAPPA), Label-free nanotechnologies in proteomics, Surface Plasmon Resonance (SPR); Modificomics: understanding post-translational modifications; Structural proteomics; Bioinformatics in proteomics; Challenges and future prospects of proteomics research.

COURSE DETAIL

Module No.	Topic/s
1	<p>An introduction to proteomics</p> <p>1. Protein structure and function</p> <p>Proteins are large organic compounds made of linear chain of monomers, known as amino acids, and are folded into a variety of complex three dimensional structures.</p> <p>Proteins are joined together by peptide bonds between the carboxyl and amino groups of adjacent amino acid residues.</p> <p>The one-dimensional information contained in the primary amino acid sequence of cellular proteins is sufficient to guide a protein into its three-dimensional structure, to determine its specificity for interaction with other molecules, to determine its ability to function as an enzyme, and to set its stability.</p> <p>1.1 Amino acids and their properties</p> <p>1.2 Amino acids form polypeptides</p> <p>1.3 Protein structure – four levels of organization</p> <p>1.4 Cellular functions performed by proteins</p> <p>2. An overview of systems biology</p> <p>The emergence of systems biology is bringing forth new set of challenges to advance biological research. Defining ways to study biological systems on a global level, integration of large and disparate data types, and issues with the infrastructural changes necessary to carry out systems biology, are various</p>



NP-TEL

NPTEL

<http://nptel.iitm.ac.in>

Biotechnology

Pre-requisites:

Basic biochemistry

Additional Reading:

Proteomics: A Cold Spring Harbor Laboratory Course Manual, A.J. Link and J. LaBaer, Cold Spring Harbor Laboratory Press, 2009.

Coordinators:

Prof. Sanjeeva Srivastava
School of Biosciences and
Bioengineering IIT Bombay

extraordinary tasks of this discipline.

Despite these challenges, the advantageous of systems biology will be far-reaching, and significant progress has already been made.

It is becoming increasingly apparent that the field of systems biology and one of its important disciplines, proteomics, can be very useful to understand complex signaling networks in biological systems.

2.1 Emergence of systems biology

2.2 What is systems biology approach?

3. Evolution from protein chemistry to proteomics

Proteomics research is grounded in classical protein chemistry and embraces the new, more complex approaches that involve high-throughput automated techniques.

Many of the techniques used under modern proteomics banner (e.g. 2-DE, MS) have actually originated several years ago from protein chemistry but now they have been adapted and after significant technological advancement become more robust and reproducible.

Proteome is defined as the total protein content of a genome and proteomics is a branch of functional genomics whose goal is to decipher the structure and function of all the proteins of a given cell at a given time under specific conditions.

The field of proteomics includes the study of various aspects of protein structure and function including expression levels, three-dimensional structure, post-translational modifications and interactions with other biomolecules to obtain a global view of cellular processes at the protein level.

Information generated from genome sequence projects, advancement of various separation techniques, rapid evolution of MS for protein identification, advent of protein microarrays for protein interaction studies, and many other novel techniques in engineering have championed the spectrum of potential of proteomics in the past few years and will lead to new insights of the biological systems in the near future.

3.1 Evolution of proteomics from protein chemistry

3.2 Proteome and proteomics

3.3 Why proteomics?

3.4 Promises of proteomics

3.5 Techniques commonly used for proteome analysis

2 Abundance-based proteomics

Proteomics based approaches provide several novel possibilities to address biological questions. Abundance-based proteomics aims to measure the abundance of protein expression by comparing control and experimental samples.

This classical proteomics approach utilizes gel-based two-dimensional electrophoresis (2-DE) and analytical mass spectrometry (MS) techniques, and present a very strong platform for the analysis and identification of proteins; however, due to many inherent drawbacks associated with the gel-based techniques, in recent years many gel-free proteomic techniques have also been developed.

Advancement of various prefractionation techniques, rapid evolution of MS for protein identification, and advent of several

gel-based and gel-free techniques have widened the spectrum of potential of proteomics to discover novel therapeutic targets and provide new insights of the biological systems.

4. Sample preparation and prefractionation steps

The first requirement for proteome analysis is the preparation of complex mixtures containing as many as several thousand proteins obtained from whole cells, tissue or organisms.

There are two steps in sample preparation – protein extraction from the source material, and solubilization of the proteins before analysis. For any proteomic study regardless of whether a chromatographic or 2D gel electrophoretic separation is performed, a critical step is the extraction and solubilization of all components.

In addition to protein extraction, it must also be soluble and free from other interacting partners (e.g. protein–RNA/DNA and protein–protein interactions).

There are various approaches reported in the literature, and many extraction procedures are now available; however, it should be noted that currently there is no single solubilization procedure that can work perfectly for all the samples.

Therefore, each laboratory must optimize protein extraction procedures for each type of tissue sample in order to get the results that best meet their experimental objectives.

4.1 Extraction and solubilization of proteins

4.2 Challenges associated with low- and high-abundant proteins

4.3 Sample prefractionation and analysis

5. Gel-based proteomics

Since inception of two-dimensional electrophoresis (2-DE) in the mid 1970s, this technique remains the core technology of choice for separating complex protein mixtures in the majority of proteome projects.

This is due to its unrivalled power to simultaneously separate thousands of proteins, the subsequent high-sensitivity visualization of the resulting differentially regulated proteins, and their relative ease.

Advancement in solubilization of hydrophobic proteins, development of immobilized pH gradient strips, gel casting and electrophoretic apparatus that can accommodate many gels minimizing the variability observed between separations, staining of gels and image analysis have furthered the development of this technique.

Two-dimensional fluorescence difference in-gel electrophoresis (DIGE) is a new development in protein detection for two-dimensional gels that allows accurate quantification with statistical confidence, while minimizing non-biological variation.

It also increases the dynamic range and sensitivity of traditional 2-DE. Various staining procedures for visualization of spots and software to compare images from different sets of gels play a crucial role in order to detect differences in protein levels between two or more samples.

5.1 Two dimensional gel electrophoresis (2-DE)

5.1.1 Staining procedures to visualize 2-D gels

5.1.2 Tools for analysis of gels

5.2 Fluorescence 2-D Difference Gel Electrophoresis (DIGE)

5.3 Blue native PAGE (BN-PAGE)

5.4 Modifications in gel-electrophoresis technique

5.5 Molecular scanner

5.6 Application of 2-DE and DIGE techniques in biological systems

5.7 Merits and demerits of gel-based proteomic techniques

6. Gel-free proteomics

Profiling the proteome of model organisms and quantitative measurement of protein expression in cells and tissue under different experimental conditions is a major goal of proteomics, for which several gel-free high-throughput screening technologies are equally available.

Gel-free techniques include multidimensional protein identification technology (MudPIT); and various techniques for functional analysis of proteome and protein-protein interactions such as protein microarrays. Another alternative technique to gel electrophoresis is MudPIT, which allows analysis of complex protein mixtures.

In this approach protein samples are subject to sequence-specific enzymatic digestion and the resultant peptide mixtures are separated by strong cation exchange (SCX) & reversed phase (RP) high performance liquid chromatography (HPLC).

Peptides from the RP column enter the mass spectrometer and MS data is used to search the protein databases. In gel-free proteomics, a combination of HPLC, liquid phase isoelectric focusing and capillary electrophoresis provides other multi-modular options for the separation of complex protein mixtures.

6.1 Two dimensional liquid chromatography

6.2 Multi-dimensional liquid chromatography based separations

6.2.1 Multidimensional Protein Identification Technology (MudPIT)

6.3 Isotope based techniques

6.3.1 Isotope-Coded Protein Label (ICPL)

6.3.2 COmbined FRActional Diagonal Chromatography (COFRADIC)

6.4 Application of gel-free techniques in biological systems

6.5 Merits and demerits of gel-free proteomic techniques

3 Central role of mass spectrometry

Mass spectrometry (MS) is preferred way for protein characterization. Before the advent of MS, protein sequences were determined by chemical or enzymatic methods like Edman degradation or amino acid analysis; however, during last decade significant improvement in chemical degradation methods, speed, automation, and higher sensitivity have rapidly advanced the throughput of MS.

Regardless of the choice of a given proteomic separation technique, gel-based or gel-free, MS is always the primary tool for protein identification & characterization, and it plays central role in proteomics studies.

7. Mass Spectrometry

Mass spectrometers consist of three basic components: an ionization source, which creates ions from the protein samples; a mass analyzer, which resolves the ions by their mass-to-charge (m/z) ratio; and an ion detector, which determines the mass of ions. In gel-free approaches such as ICAT and MudPIT, samples are directly analyzed by MS; whereas, in gel-based proteomics (2-DE and 2-D DIGE), the protein spots are first excised from the gel and proteins are digested with trypsin.

The resulting peptides are then separated by LC or directly analyzed by MS. The development of specialized ionization techniques such as Matrix Assisted Laser Desorption and Ionization (MALDI) and Electrospray Ionization (ESI) as well as design and development of new types of mass spectrometers with improved sensitivity, selectivity and mass accuracy have significantly improved the quality of data generated in proteomic experiments.

7.1 Ionization source

7.2 Mass analyzer

7.3 Ion detector

7.4 Different types of mass spectrometers and modifications

7.5 Mass spectrometry applications

7.6 Merits and demerits of different types of mass spectrometers

8. Mass spectrometry data analysis – computational tools

The accuracy, high throughput, and robustness of MS technologies, have made the characterization of entire proteome a realistic goal.

The experimental peptide masses are correlated to the peptide fingerprints of known proteins in the databases through search engines such as Mascot and Sequest. The tandem spectra of amino acids are used to search the databases.

8.1 Mass spectrometry data analysis

8.2 Search engines for MS protein identification

4

Quantitative proteomics

9. Quantitative proteomics

Profiling the proteome of model organisms and quantitative measurement of protein expression in cells and tissue under different experimental conditions is a major goal of proteomics, for which several gel-free high-throughput screening technologies are equally available.

Gel-free techniques include isotope-coded affinity tag (ICAT); isobaric tagging for relative and absolute quantitation (iTRAQ); multidimensional protein identification technology (MudPIT); and various techniques for functional analysis of proteome and protein-protein interactions such as protein microarrays.

In order to increase the quantitative measurement of proteins ICAT technique enables isotope tagging of specific proteins chemically in two separate samples and gives a quantitative measure of protein expression changes. The iTRAQ technology, a variation of ICAT, is similar in concept. ICAT relies on tagging cysteine residues; whereas, in iTRAQ method tagging is on primary amines.

Another alternative technique to gel electrophoresis is MudPIT, which allows analysis of complex protein mixtures. In this approach protein samples are subject to sequence-specific

enzymatic digestion and the resultant peptide mixtures are separated by strong cation exchange (SCX) & reversed phase (RP) high performance liquid chromatography (HPLC). Peptides from the RP column enter the mass spectrometer and MS data is used to search the protein databases.

In gel-free proteomics, a combination of HPLC, liquid phase isoelectric focusing and capillary electrophoresis provides other multi-modular options for the separation of complex protein mixtures.

9.1 Gel-based quantitative proteomics

9.1.1 Fluorescence 2-D Difference Gel Electrophoresis (DIGE)

9.2 Gel-free mass spectrometry based quantitative proteomics

9.2.1 Stable Isotope Labeling by Amino acids in Cell culture (SILAC)

9.2.2 Isotope Coded Affinity Tag (ICAT)

9.2.3 Isobaric Tagging for Relative and Absolute Quantitation (iTRAQ)

9.2.4 Proteolytic labeling with [18O]-water

9.3 Application of quantitative proteomics

9.4 Merits and demerits of gel-free quantitative proteomic techniques

5 Functional proteomics

Functional proteomics aims to determine the role of proteins by assessing protein interaction, biochemical activities and it requires high-throughput tools to elucidate the cellular networks & interactions between proteins. In functional proteomics research one of the major goals is to study the circuits of protein interaction networks that regulate the lives of cells and organisms.

Traditional biochemical methods such as yeast two-hybrid, immunoprecipitation, and advent of microarray technology have advanced the interactomics research.

In order to understand the network of responses in proteomic analysis integration of new label-free and nanotechniques such as Surface Plasmon Resonance (SPR), Atomic Force Microscopy (AFM), carbon nanotubes & nanowires have potential for real-time monitoring of interactions.

Another level of protein complexity arises from numerous co- and post-translational modifications which generates tremendous heterogeneity and extreme diversity of physico-chemical properties of proteins.

PTMs such as glycosylation, phosphorylation and proteolytic processing play an essential role in the function of the protein.

10. Interactomics: techniques to study protein-protein interactions

Many proteins recognize and bind to other proteins. The branch of proteomics which aims to study protein-protein interactions is referred as "interactomics".

Traditional biochemical methods such as yeast two-hybrid and immunoprecipitation (IP) followed by MS have been most extensively used to study the protein-protein interactions; however, protein microarrays offer many advantages including high-throughput analysis of thousands of proteins

simultaneously.

The early applications of microarrays have focused on DNA-based applications by using DNA microarrays to profile the gene expression changes; however, these arrays cannot provide information about protein PTMs or protein-protein interactions.

Due to the central role played by the "verbs of cell", the protein, several approaches have been used to generate protein microarrays. Protein microarrays consist of numerous capture agents such as proteins, antibodies and peptides that selectively bind target proteins to a chip surface and provide a high-throughput platform for the structural and functional characterization of large numbers of proteins.

10.1 Yeast Two-Hybrid (YTH)

10.2 Immunoprecipitation (IP)

10.3 Protein microarrays

10.3.1 Abundance-based microarrays

- a. Capture microarrays
- b. Reverse-Phase Protein (RPP) microarrays
- c. Tissue microarray (TMA)

10.3.2 Function-based microarrays

- a. Chemically linked microarray
- b. Peptide fusion tags
- c. Protein microarrays by using cell-free expression system

Nucleic Acid Programmable Protein Array

(NAPPA) Protein in situ array (PISA)

Multiple spotting technique (MIST)

10.3.3 Other array formats

- a. Microwell arrays
- b. Microfluidic chips
- c. Cell-based arrays
- d. Small Molecule Microarrays (SMMs)

10.4 Protein-protein interactions to understand biological systems

10.5 Pros and cons of using various interactomics techniques

11. Label-free nanotechnologies in proteomics

In proteomics studies the labelling strategies have synthetic challenges, multiple label issues, and they exhibit interference with the binding site therefore, a variety of methodologies and nanotechnologies such as Surface Plasmon Resonance (SPR), Atomic Force Microscopy (AFM), carbon nanotubes & nanowires, MEMS cantilevers, and SELDI are in various stages of development.

Label-free technologies make it possible to measure biomolecular interactions in real-time with high degree of sensitivity.

11.1. Nanotechnologies in proteomics

- 11.2. Surface Plasmon Resonance (SPR)
- 11.3. Atomic Force Microscopy (AFM)
- 11.4. Carbon nanotubes and nanowires
- 11.5. Electrochemical Impedance Spectroscopy (EIS)
- 11.6. Application of nanotechnologies in proteomics
- 11.7. Nanotechnologies in proteomics: merits and demerits

12. Modificomics: understanding post-translational modifications by using proteomic techniques

Post-translational modifications (PTM) of proteins play a crucial role in cellular function. It generates tremendous diversity, complexity and heterogeneity of gene products and its identification can be quite challenging.

An important goal of proteomics is the identification of PTMs. The area of proteomics which deals with PTMs is known as “modificomics”. PTMs can turn the 20 specific coded-encoded amino acids into more than 100 variant amino acids with new properties.

Proteins can undergo a range of PTMs such as phosphorylation, glycosylation, sulphonation, palmitoylation and ADP-ribosylation. These PTMs, with a number of other modifications, can considerably increase the information content and the functional repertoire of proteins.

Functional consequences of PTM can range from preservation of molecular structure to verify dynamics and flexible changes in the functional activities of the proteins.

Detection of PTMs usually requires careful analysis by mass spectrometry. Few staining techniques have also exhibited promise for PTMs analysis.

- 12.1 What are post-translational modifications?
- 12.2 PTMs pose challenge for proteomics and bioinformatics
- 12.3 Techniques for characterization of PTMs
 - 12.3.1 Gel electrophoresis and staining procedures for PTM identification
 - 12.3.2 Identification and quantitation of PTMs by MS

6

Structural proteomics

Completion of genome sequencing projects have provided us with wealth of information in the form of gene sequences; however, these sequences must be related to the proteins they encode and in-turn their biological and biochemical significance.

The genome-wide approach to protein structure determination, termed “structural proteomics”, provides a new rationale for structural biology. Traditionally, structural biologists attacked a problem only after it had been firmly characterized using biochemical and/or genetic methods.

However, relying on structure-function relationships, it will now be possible to suggest a biochemical function of uncharacterized proteins based solely on structural homology to another protein with a known function.

Such a predicted function could then provide the foundation for a hypothesis that could be tested with additional biochemical experiments.

The 3D structure of a protein polypeptide chain determines its biochemical function; the building of structure-function correlations for novel and diverse protein conformations is next step in structural proteomics.

Identification of all the proteins on the genome-wide scale, determination of their structure-function relationships, and outlining their precise 3D structure are major challenges in structural proteomics.

13. Protein purification: affinity-based, cell-free and high-throughput approaches

Almost all proteins subjected to structural studies are expressed in heterologous systems.

Recent developments in molecular biology have allowed high-throughput cloning and expression of proteins. However, overexpression in host cells may result in inappropriate PTM and incorrect folding, which usually forms inclusion bodies and insoluble aggregates, and thus must be discarded.

Such limitations have been conquered by a number of strategies, such as using genes from different species, altering constructs, screening for solubility, and utilizing different cellular or cell-free expression systems.

Furthermore several novel refolding techniques such as refolding chromatography based on molecular chaperones that are part of the cellular machinery responsible for folding nascent protein and the designed small-molecule agents to assist the refolding process, have been developed to recover these proteins.

Various protein purification techniques including nickel nitrilotriacetic acid (Ni-NTA) agarose affinity chromatography, cell-free protein production and high-throughput protein production will be described.

13.1 Tag-based protein purification

13.2 Cell-free protein production

13.3 High-throughput protein purification

14. Structural proteomics

To help annotate the structure and biochemical function of proteins on a genome-wide scale several techniques such as X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and computational methods such as comparative and de novo approaches, molecular dynamic simulations, are intensively used.

Historically, structure determination and the characterization of dynamic processes by x-ray diffraction and NMR have lagged behind the genetic and biochemical characterization of cellular processes.

However, new advances in x-ray diffraction (e.g. use of robotics to ensure successful growth of crystals, the development of synchrotron sources for macromolecular crystallography, the use of direct methods and anomalous dispersion in phasing, and the development of automated methods of map fitting) and NMR (e.g. availability of higher magnetic field strengths, development of labeling schemes and relaxation-compensated experimental methods to alleviate the inherent size limitation associated with NMR, and the use of residual dipolar coupling to obtain important long-range structure constraints) have made structural proteomics a reality, by permitting the atomic level description of the structure and dynamics of a large number of complex biological assemblies.

	<p>14.1 X-ray crystallography</p> <p>14.2 Nuclear Magnetic Resonance (NMR)</p> <p>14.3 Computational methods</p> <p>14.4 Structure prediction from sequence</p> <p>14.5 Deriving function from sequence</p> <p>14.6 Application of structural proteomics</p> <p>14.7 Merits and demerits of structural proteomics techniques</p>
7	<p>Bioinformatics and proteomics: computational models of proteomic networks</p> <p>15. Informatics in proteomics</p> <p>The high-throughput proteomic technologies are generating huge amount of data and its handling and analysis represent challenge for the scientific community.</p> <p>New collaborations between proteomics scientists, bioinformaticians and biostatisticians are now emerging to develop robust, sensitive, and specific methodologies & tools for the analysis of proteome.</p> <p>There is need to develop efficient and valid methods of data analysis; develop variety of different databases; develop tools to translate raw data into the forms suitable for data analysis; and develop user interfaces to visualize data.</p> <p>15.1 Bioinformatics and proteomic technologies</p> <p>15.2 Public protein databases and interfaces</p> <p>15.3 Data mining in proteomics</p> <p>16. Modelling of proteomic networks</p> <p>Scientists in proteomics research and computer science are collaborating, and several efforts are currently underway to analyze and interpret experimental data, generate models for determining ontologies and modelling protein interaction networks. Directions for on-going and future work to model proteomic networks will be discussed.</p> <p>16.1 Need for model prediction</p> <p>16.2 Generation of models in proteomic studies</p> <p>16.3 Current endeavors and future challenges</p>
8	<p>Functional proteomics</p> <p>17. Challenges and future prospects of proteomic research</p> <p>Genomics represent only the first step towards an understanding of cellular and higher order functions and it has to be complemented by the systematic analysis of the proteins i.e. proteomics.</p> <p>Although the function of a large number of proteins encoded by genes in the genome are known, we are still far from understanding those of an equally large number of the remaining proteins specified by other genes.</p> <p>Proteomics has emerged as an indispensable tool for understanding cellular mechanisms, and its scope in biological research are much broader than it was originally realized.</p>



From the initial objective of proteomics to identify as many individual proteins as possible in a given biological sample to the development of high-throughput, parallel and quantitative technologies for analyzing proteome dynamics, the scale and focus of proteomics research is now shifting towards functional analysis.

The tremendous heterogeneity, numerous post-translational modifications and protein-protein interactions pose several challenges in proteomics research.

New proteomics methodologies, together with other computational tools, provide a platform to study regulatory networks underlying the function of living organisms and they can help scientists to analyze cellular functions with speed, reproducibility and accuracy.

References:

1. Introduction to Proteomics: Tools for the New Biology, D.C. Liebler, Humana Press, 2002.
2. Principles of Proteomics, R.M. Twyman, Bios Scientific Pub., 2004.
3. Proteomics for Biological Discovery, T.D. Veenstra, J.R. Yates III, John-Wiley & Sons, Hoboken, New Jersey, USA; 2006.
4. Protein Biochemistry and Proteomics (The Experimenter Series), R. Hubert, Academic Press, 2006.
5. Proteomics in Practice: A Guide to Successful Experimental Design, R. Westermeier, T. Naven, H-R. Höpker, Wiley-VCH, 2008.